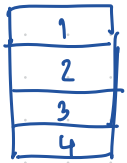
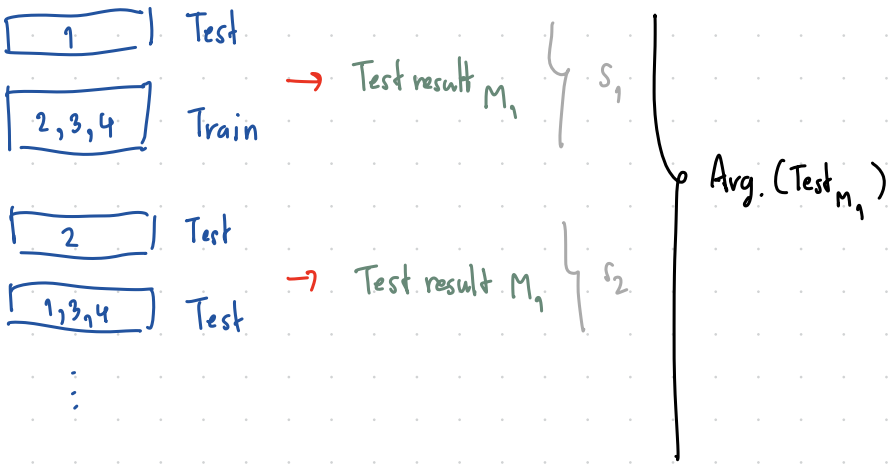
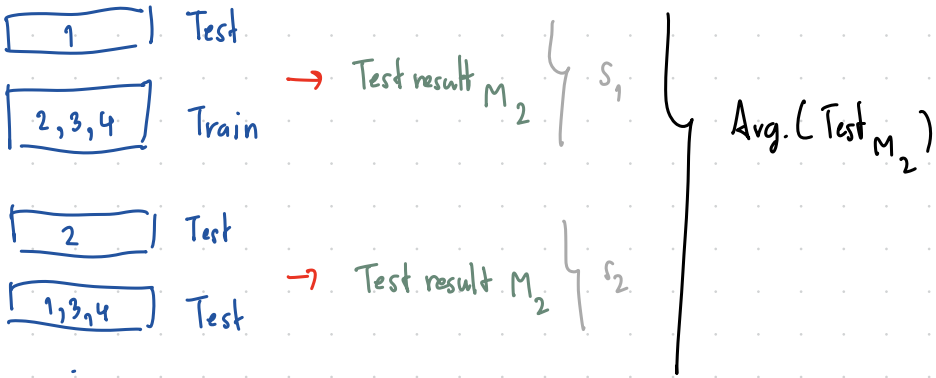
Cross-Validation concept



Model ที่ 1 (M_1)



Model ที่ 2 (M_2)



∴ เราจะเลือก Model จาก ค่า Avg.(Test) ตัวไหนมีมากกว่า

พอได้ Model ไหนดีกว่า

$M_1 \rightarrow Avg.(Test)_{M_1}$

$M_2 \rightarrow Avg.(Test)_{M_2}$

M_2 ดีกว่า $\rightarrow$

Test กับ data รวม

1
2
3
4

$\downarrow$

$\overline{M}$ model ที่ train แล้ว

Test จากการแบ่ง folds

1
2
3
4

$\therefore$ พยายามทำให้ distribution ของข้อมูล ใน train / test เหมือนกัน